

State of Michigan



Document Imaging Guidelines

“Responsibility of Source Document Owners to Monitor the Work”

With limited human resources, the time it takes to maintain, process, and store records is becoming more challenging each day. New technologies are available to help. But the challenge is not the technology, it is the use and integration of the technology into your daily routine.

The creation of records is a fundamental aspect of the management of any business operation, government or private. Information is an asset. Therefore, it is important that agencies determine how and why records are being created. Consideration should be given to capturing appropriate content, and design considerations at the beginning of record creation.

Due to legal requirements, the management of records and information is critical to state government. The institutional knowledge base of the organization depends on its ability to store and access relevant information.

Technologies for capturing, storing, retaining, and accessing information are expanding the volume of information that can be acquired, with increased longevity.

Records management is changing from the traditional focus on the form of records and their storage, to a broader focus on how a record's content and value relate to legal issues and business needs. Using information to create value and plan strategically is a driving force in today's world. Preparing documents in anticipation of the need to store and retrieve quickly in a work environment, with reduced human resource numbers must be a key part of the State of Michigan's strategic efforts for all departments and agencies.

This booklet provides valuable information to help agencies to plan, understand and streamline their efforts to use the latest imaging technology.

The booklet focuses on Document Imaging.

Statement of Work

The first step in document imaging project is the development of a statement of work. The statement contains the general and specific tasks and responsibilities of the imaging contractor and the State of Michigan. It clearly lays out the scope and purpose of the imaging project as it pertains to document preparation, document indexing, quality control and quality assurance.



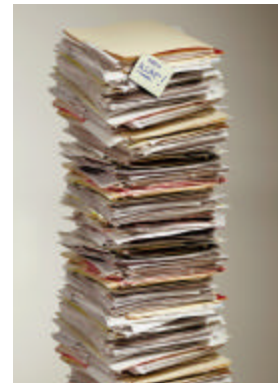
The following sections highlight those areas most important for your agency to address to ensure an eye-readable end product using imaging. These are:

- Document Preparation
- Document Indexing
- Quality Control
- Quality Assurance

Document Preparation

Document Preparation is important to ensure accuracy and quality of final product. A document preparation process plan must be defined, developed and recorded in detail for each document type that will be scanned.

The focus of this plan should zero-in on “Back- File” documents, which are those documents currently stored; and “Day-Forward” documents, which are those documents currently in use and needed day to day.



Document Preparation is very labor-intensive. Someone from the agency who is considered to be the “document expert” must be involved in this process. This person must have expert knowledge of the relevant files and documents, and be able to guide the effort. This is very important in order to determine accurate document indexing, to ensure the right document look when accessed through an Electronic Document Management System.



Document preparation design identifies the pages of documents that will be scanned. All document versions will be categorized. Every document in the physical file, whether converted or not, must be addressed. Most importantly, the outcome of this effort will determine the order of each document type and how it appears in the scanning process. This must be detailed in a written conversion process plan.

Procedures are then developed that include instructions for extraction of staples, use of paper clips, removal of post-it notes and when to copy a new page in order to enhance or ensure clarity of the source document. The procedures must also address varying paper sizes and weights, and the repair of torn pages.

Assistance in this effort is available through Records Management Services Analyst staff. An outside vendor, under the division’s direction, will guide your efforts in the preparation process.

Document Indexing

Indexing is the key to effective imaging product utilization. It is through indexing that you will make use of the image documents. Indexing gives your records value in the daily operation of your agency. It is how you take your scanned conversions of documents, film, books or old computer files and turn them into organized, and useful digital information.

There are four indexing levels to consider:

1. Folder
2. Sub-Class
3. Document Level
4. Full Text Search

Documents

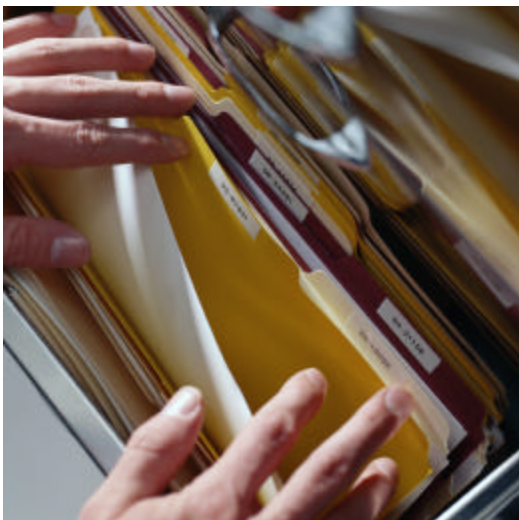
"A container of information...organized and presented to provide meaning and context to the information...it can be a form, letter, spreadsheet, memorandum, photograph, video clip, or report."

Folder is the largest level breakdown. For example, a group of similar personnel-related documents are placed in a folder by calendar year.

Subclass is the next level down. For example, a group of similar personnel-related documents are placed in a folder by calendar year and month.

Document Level continues to drill this down further. For example, a group of similar personnel-related documents are placed in a folder by calendar year, then month, then employee name.

Full Text Search takes indexing to the lowest level in this example. A group of similar personnel-related documents are placed in a folder, by calendar year, then month, then employee name and finally through "keyword" search of actual document. In this example the "keyword" search could be to determine if specific training was received.



Documents are imaged based on the criteria and specifications established by the user.

For example you may have:

- Capability to manually enter a search by index fields. Something not available with a paper filing system
- Second Pass entry and verification of critical index fields can be used for almost perfect accuracy.
- Full Text indexing allowing searches for document images by their content.

After the document index process is complete, retrieval is quite simple.

It is important to keep in mind that there is no specific right or wrong way to set up an indexing structure. Each

has advantages and disadvantages and must be assessed for their value to the system being implemented, and their costs.

Document Scanning

Document scanning uses document-scanning software and high-speed scanning hardware to convert a paper or microfilm image to a digital picture.

Scanning involves taking documents or microfilm and converting them to (pixel based) images in standardized file formats. For most projects, source documents are first scanned to tag image file format (tiff) or portable document format (pdf), and are saved on CD media. In researching scanning options, you will need to take into account document characteristics.

The challenge to scan documents in a readable format is the wide range of document types. Various patterns of content can exist in one document. These could include line art, text, or photos. The documents and content must be analyzed to determine best scanning options.



Assuming the ideal case, the general process used for selecting a scanning method is as follows:

- Examine usage of the images and information and type of documents intended for scanning.
- Examine the content present in the documents.
- Choose a scanner, output devices, and compression standards.
- Choose the options for the compression scheme. A compression scheme is a method to reduce the amount of data needed to store or transmit a representation of a specific image.

The best compression scheme to choose is determined by both the type of document and the graphical contents within it. The construction of the document is a critical consideration in determining scanning options. This includes:

- Bound or unbound documents
- Mass or hand printed documents
- Ink and paper type used
- Whether document is original or reproduction
- Continuous tone
- Black and white photos or text
- Multicolor text, line art, half tone style or full color documents

Stand-alone systems are designed to handle typical office documents. These systems are not viable options if you need to convert odd-size documents or microfilm. Your agency may already be in the business of converting some or all of your documents using scanning equipment. You may have already acquired an imaging system that includes scanning, storage and some image retrieval capabilities. This can be integrated with additional technologies.

The issues of imaging include: knowing when and why scanning documents makes sense. Scanning paper-source documents can save storage space that translates to dollars saved. Using an estimate of 2,500 pages per file drawer and four file drawers per file cabinet, it is easy to do the math. Scanning one filled four-drawer filing cabinet full of documents total 10,000 single sided pages. That will fill one CD-ROM disk.

Similarly, the scanned contents of two file cabinets will fill one gigabyte (GB) of magnetic disk storage. The point here is to recognize advantages of scanning in reducing space requirements to house your existing paper-source documents and records.

Quality Control at Production

Record keeping requirements must include emphasis on quality source documents. The agency or department, as document owner, must determine the level and degree of inspection necessary to ensure that acceptable quality levels are met. The contractor charged with responsibility to convert documents to acceptable images must establish appropriate quality control measures. These should include at minimum:



- Conversion application procedures to ensure that staff involved in the process have the training and time to perform the tasks necessary for a quality output.
- Develop conversion systems to allow for both quantity and quality measures for each task in the conversion process.
- Continually monitor the document conversion process while in production, always looking for better ways to do the job.

Active agency/department participation through the entire process is important. At the end when the scanned product is delivered, quality assurance rests with the source-document owner.

Quality Assurance at Delivery

The purpose of quality control is to establish a standardized approach for the systematic, statistically valid verification of work performed by the imaging contractor. Quality assurance is met in partnership with Records Management Services and the agency/department-owner of the source document. Upon receipt of an imaged document from the contractor, the agency must do the following to ensure accuracy, quality and timely storage:

- Check to make sure imaged document matches request
- Return source document to Records Management Services for proper storage if a back-file product
- Check duplicate image document to ensure quality

The agency's role is to monitor the accuracy of the imaged document. Verification is needed to ensure that any information is readable, can be recognized in the original, and can also be recognized on the digital images created.

Electronic Imaging

"The process of capturing, storing, and retrieving documents that were once hard copy (paper or microfilm) using bit mapping....converting hard copy to digital images."

As the end user you must understand your role to ensure quality. The recommended approach is designed to provide a proven method of validating the overall quality of the delivered work product.

Keep in mind that there is a significant difference between the quality control steps provided by the imaging vendor, that are designed to detect and correct errors in the production process, and quality assurance which is designed to verify the validity and accuracy of the overall delivered product.

While the imaging vendor will provide quality control prior to product delivery, the end user must also perform its own quality assurance in order to verify that the delivered work product is acceptable. The end user quality assurance process must be performed in a timely manner in order to conform within the agreed upon acceptance terms (usually 30 days) of the existing contract.

Sample for Quality

In order to establish a meaningful sampling process there are three categories of information that must be established. The end user must:

- identify the specific attributes of the work product that are critical to them,
- establish the acceptability level expressed as a percentage,
- establish a batch size expressed as a number of items that are contained within the batch.

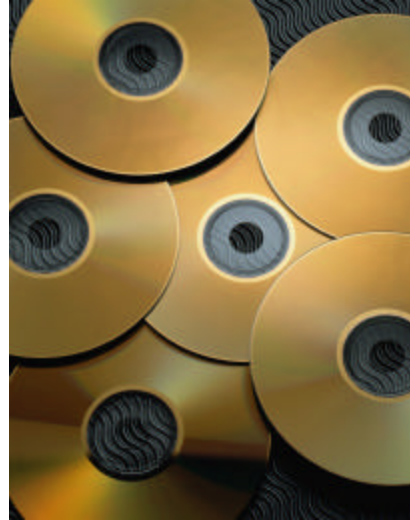
From this information an inspection model can be developed that will, within the limits of the acceptability level, assure the end user that the delivered work product has met the established standard.

Attributes

The specific attributes that need to be defined are those component elements of the imaging process that have been determined to be critical to the overall success of the conversion process. In the vernacular, it is not sufficient to simply say, “this is a good image.” Objective criteria that define what a good image is must be established so that the production process can routinely and reliably produce the defined “good image.”

Common examples of attributes are:

- A. Scanning Resolution
- B. Image Legibility
- C. Image Skew
- D. Image Rotation
- E. Image Cropping
- F. Index Data Accuracy
- G. Image and Index format compliance with the system upload requirements.
- H. Document Type accuracy
- I. Image File Format
- J. Data Format



The end user may also identify other attributes. The above examples are not intended to be an exhaustive, totally comprehensive list. As the end user you must be able and willing to identify those elements that are considered important in the identification of a "good image."

Remember that these attributes must be defined objectively, because subjective attributes are not reliably measurable. Be careful to select attributes that are critical to you, and do not select attributes that are not critical to you. By way of example, if a border around the image is acceptable to you, do not select it as an attribute if it will cause additional cost. If elimination of all borders is critical to the success of your system, then define elimination of borders as an attribute, even if there is an additional cost associated with the elimination of borders.

Acceptability Level

For each attribute selected and defined, a level of acceptability must be established. This level is expressed as a percentage. The higher the acceptability levels, the greater the number of items that must be inspected within a batch to ensure that the work product meets the established expectations.

Select an acceptability percentage for each attribute that provides adequate protection against errors. More critical attributes should be assigned higher levels than less critical attributes. A 100% certainty is not a realistically achievable level therefore 100% cannot be selected as the acceptability level. It is important to understand that the higher the selected acceptability level the greater the work effort required to achieve it and generally it will result in an increase in cost. By way of example, if the accuracy of the index data is highly critical, as it most certainly should be, then you might set the acceptance level at a high level, such as 99.5%

Batch Size

This is critical because it will ultimately determine the number of items that need to be inspected within the batch. It is generally acceptable to employ natural batches for control purposes. For example, if documents are managed as files, a natural batch may be a single file or a multiple of number of files. If natural batches are not readily available, then a batch can be established by counting it out. You would then isolate the desired number of items that become the batch.

From the above information a sampling plan is established using American National Standards Institute, ANSI/ASQC 21.9 and ANSI/ASQC 21.4 standards. These established and uniformly accepted standards are used to evaluate the overall quality of delivered work products by companies who provide goods and services to commercial enterprises, and governmental departments and agencies.

Proper design, implementation and execution of the quality assurance process will provide you with the knowledge that the images and data in your system meet your expectations. Records Management Services can assist you in the design of a quality assurance process that will meet your needs and provide the protection against errors that you expect.

Electronic Document Management System

“a set of computer- based technologies which allow the electronic capture, processing, indexing and distribution of documents as graphical images, as well as text based images, in order to manage document- based information.”

Frequently Asked Questions

How does an agency or department determine or prioritize what documents need to be stored for retrieval and what documents can be tossed?

Agencies and departments need to develop a records retention plan in conjunction with Records Management Services. The plan establishes a partnership between the agency/department and Records Management in the proper storage and disposal of state records. Retention schedules establish retention periods that are based on the business needs of the organization.

What is Document Imaging and when should my agency/department consider this option for document management?

Document imaging is the process of converting paper or microfilm documents to digital images accessible via your computer. The benefit of imaging is the accessibility by multiple users and the ability to quickly resume business in the case of a disaster.

The drawback to imaging is the uncertain long-term future of the media and format. In order to protect the States information resources, Michigan law currently requires any record with a retention value of greater than 10 years to also be retained and accessible in an eye-readable format (microfilm or paper).

If it is our agency's desire to image documents, what are the procedures to get this done?

Records intended for imaging must first be placed on an agency specific Records Retention and Disposal Schedule, or be covered by one of the General Schedules. Contact Records Management Services at 517-335-9132 for information and guidance to pursue development of a document imaging application.

Records Management Services will assist you in justifying the need for the imaging application, include the application on your Retention and Disposal Schedule, and assist you in preparing for production. This service is free to all state agencies. Costs of conversion services will vary and be dependent on the requirements of each application.

What are the most critical considerations to be determined in the development of a document imaging system?

When developing an imaging system, key consideration should be given to development of a technology strategy that includes the following:

- Preservation of current imaged records on existing media with attention given to environmental storage;
- Maintain existing hardware and software through timely upgrades;
- Develop a plan to transfer images, index and data as successive versions of hardware and software become available; and
- Migrate optical imaging systems as new technologies, as yet undefined, are developed.

Finally, always budget for change. The costs of hardware and software has dropped in over the years, so the real costs of an imaging system are centered on training, support, conversion of documents to digital formats, and continuous system upgrades needed to stay current with the latest technology. Annually budget between 10 to 20% of the cost of the original system for maintenance, support and upgrade.

What are the advantages of this technology?

The advantages and benefits are many. These include reducing the physical storage space needed to house your records. For example up to 50,000 pages of documents can be stored on one CD-ROM.

Documents that are imaged and are part of an electronic document management system can be retrieved quickly, allowing documents to be shared across a network. A key component of any document management system is the ability to index and cross-reference documents. The indexing process allows for searches on keywords, version control to keep track of progressions in the development of a finished document, a check-in /check-out process to monitor documents removed and returned, and a security function to determine authorization level to review documents.

What types of documents can be scanned?

There is virtually no limit to the type of document that can be scanned. The size can range from a small receipt to as large as an E-size technical drawing. "Documents" also describes pictures, posters, brochures, etc. all of which can be scanned.

What is scanning software?

Scanning software is a tool that enables the computer to take a digital image (or picture) of the document in question. This software identifies optimums in shading, clarity, contrast and brightness enabling the hardware to make appropriate adjustments during processing to ensure the most crisp image display possible.

How do I make my documents "Scan Ready"?

"Scan ready" means that the paper document is "prepared" to go through the scanning equipment. Preparing documents requires that all staples have been removed from the files and the pages are not excessively wrinkled, torn or otherwise damaged. The pages are all oriented the same way (i.e.: all tops up) and facing the same way (if single-sided). The files are separated and labeled in a way that is logical to how you would store and access them normally. Sometimes sorting may also be desirable. Records Management Services can assist you in determining the best labeling and indexing scheme for your needs.



Where can I get more information and assistance on developing an imaging system?

The Department of History, Arts and Libraries, Records Management Services provides technical advice and assistance to agencies. A consultant can meet to discuss your needs and guide your agency to the best solution. For more information contact: Michigan Department of History, Arts and Libraries , Records Management Services, 3405 North Martin Luther King Blvd., Lansing, MI 48909, Phone: 517-335-9132 Fax: 517-335-9418 Email: recordscenter@michigan.gov

Glossary of Terms

Automated Retrieval – Retrieval system in which the images are displayed automatically. Commonly the user interrogates an index, which may be manipulated by a computer to locate the images.

Back-file Conversion – Process of scanning, indexing, and inspecting a large existing collection of documents.

Bit Map – Method of representing images by assigning a block of memory that stores pixel images in a device specific format

Compact Disc-recordable (CD-R) – Optical disc in which data is recorded once by the user and can be read many times.

Conversion – Process of changing a physical document to another form, e.g. paper to digital image, microfilm to digital image.

Data – Representation of facts, documents or instructions in a formalized manner suitable for communication, interpretation, or processing by humans or computers.

Digital – Use of binary code to record information. Information can be text in a binary code, or images in a bit-mapped, or sound in a sample digital form or video.

Electronic Image – A digital representation of a document.

Enhancement – Technique for processing an image so that the result is visually clearer than the original image.

Expungement – Process of removing a document from a system and leaving no evidence of the document ever having appeared in the system.

File – An organized collection of records and information directed toward some purpose.

First Generation Image – Image, generally used as a master, produced directly from a subject.

Index – List of the contents of a file, document or collection together with keys or references for locating the contents.

Information Retrieval – Methods and procedures for recovering specific information from stored data. For example, selection of a specific document from an optical disk.

Jukebox – Automated device for housing multiple optical disks and one or more read/write drives.

Keyword – One of the significant and informative words in a title or document that describes the content of that document.

Local Area Network (LAN) – a dedicated connection between computers that allow intercommunication and sharing of resources such as disk storage and printers. LAN systems are usually confined to a small area or site: department or agency.

Parallel Processing – Simultaneous execution of two or more process operations in a single computer system.

Pixel – Smallest element of a display surface that can be independently assigned color or intensity. It is synonymous with picture element.

Portable Document Format (pdf) – File format which captures formatting information from a variety of applications and makes it possible to transmit and display documents in an identical way, independent of the platform.

Retrieval – Workflow engine function that organizes retrieval of documents from storage or other sources such as a server.

Tag Image File Format (TIFF) – Widely used image file structure that consists of a series of headers or tags, plus the image data.